

УДК 004.021

## АЛГОРИТМЫ ИНТЕГРАЦИИ РЕЗУЛЬТАТОВ РАСПОЗНАВАНИЯ В ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЯХ ПОЛЕЙ ДОКУМЕНТОВ, УДОСТОВЕРЯЮЩИХ ЛИЧНОСТЬ

<sup>1</sup>Булдакова Т.И., <sup>2</sup>Славин О.А., <sup>2</sup>Путинцев Д.Н.

<sup>1</sup>Московский физико-технический институт, Москва;

<sup>2</sup>Институт системного анализа ФИЦ ИУ РАН, Москва, e-mail: 2001dnp@mail.ru

В статье рассматривается проблема интеграции результатов распознавания отдельных полей документа с целью оптимизации быстродействия в контексте задачи распознавания документа с помощью мобильных устройств. Задача распознавания документов исследуется на примере паспорта гражданина РФ. Исходными данными служит видеопоток (набор кадров), полученный камерой мобильного устройства. В отличие от изображений, получаемых со сканера, при работе с видеопотоком документ может располагаться в некоторой произвольной плоскости относительно плоскости сфокусированного изображения, что добавляет задачу проективного преобразования полученных кадров. В задаче распознавания документа выделяют этапы детектирования документа в кадре, нахождения зоны и параметров проективного восстановления, проективное исправление изображения документа, позиционирование зон реквизитов на исправленном изображении, распознавание реквизитов, и постобработка. В работе предложен алгоритм интеграции, который позволяет получить высокую точность распознавания полей при существенном ограничении качества исходных данных для распознавания.

**Ключевые слова:** обработка изображений, распознавание, поле документа, анализ изображений, интеграция результатов

## ALGORITHMS OF INTEGRATION OF THE RESULTS OF RECOGNITION IN VIDEO SEQUENCES OF THE FIELDS OF IDENTITY DOCUMENTS

<sup>1</sup>Buldakova T.I., <sup>2</sup>Slavin O.A., <sup>2</sup>Putintsev D.N.

<sup>1</sup>Moscow Institute of Physics and Technology (State University), Moscow;

<sup>2</sup>Institute for Systems Analysis, FRC CSC RAS, Moscow, e-mail: 2001dnp@mail.ru

The article discusses the problem of integrating the recognition results of individual document fields with the purpose of optimizing performance in the context of the task of document recognition using mobile devices. The task of document recognition is investigated on the example of a Russian citizen's passport. The initial data is the video stream (a set of frames) received by the camera of the mobile device. Unlike images received from the scanner, when working with a video stream, the document can be located in some arbitrary plane with respect to the plane of the focused image, which adds the task of projective transformation of the received frames. In the task of document recognition, the document detection stages in the frame are selected, the area and the parameters of the projective recovery are found, projective correction of the document image, the positioning of the requisite zones on the corrected image, the recognition of the details, and post-processing. The paper proposes an integration algorithm that allows obtaining high accuracy of field recognition with a significant limitation of the quality of the initial data for recognition.

**Keywords:** image processing, recognition, document field, image analysis, results integration

В последние несколько лет большое внимание уделяется исследованию вопросов распознавания различных объектов, в частности, текстовых документов, с использованием мобильных телефонов, как в качестве устройств захвата изображения, так и в качестве вычислительных устройств. Так, в разных работах обычно подходят к рассмотрению каких-то подзадач, например, таких, как оценка качества изображений, отдельные задачи распознавания, постобработка полученных результатов. При этом серьезному рассмотрению процесса в целом, а особенно возникающим обратным связям уделяется значительно меньше внимания. В основном рассматриваются системы, в которых захват кадра производится мобильным телефоном, а распознавание выполняется на сер-

вере. В данной работе мы рассматриваем схему построения системы распознавания на устройстве, возникающие обратные связи и их применение для повышения качества распознавания.

Рассматривается задача распознавания документов на примере паспорта гражданина РФ с помощью мобильного устройства. Исходными данными служит видеопоток (набор кадров), полученный камерой мобильного устройства. В отличие от сканера, документ может располагаться в некоторой произвольной плоскости относительно плоскости сфокусированного изображения, что добавляет задачу проективного преобразования полученных кадров. Задача распознавания документа, как правило, разделяется на следующие этапы:

– детектирование документа в кадре;

- нахождение зоны и параметров проективного восстановления;
- проективное исправление изображения документа;
- позиционирование зон реквизитов (полей) на исправленном изображении;
- распознавание реквизитов;
- постобработка.

В случае видеопотока задача постобработки является особенно актуальной. Межкадровая информация позволяет повысить точность и скорость получения результата распознавания [2]. Мы остановимся на решении задачи интеграции результатов распознавания отдельных полей и на организации управления механизмом распознавания с целью оптимизации быстродействия.

### Описание алгоритма

Пусть для каждого отдельно взятого документа есть набор из  $n$  распознанных кадров  $I_0, I_1, \dots, I_{n-1}$  соответственно с  $m$  распознанными полями  $F_0(I_i), F_1(I_i), \dots, F_{m-1}(I_i)$ . Для простоты будем рассматривать последовательность результатов для одного конкретно взятого поля. Каждое поле представимо в виде набора знакомест:

$$A(I_i) = \{A_1(I_i), \dots, A_l(I_i)\},$$

где  $l=|I_i|$  – количество распознанных знакомест в поле (это количество зависит от кадра и равно 0, если поле не распознано вообще), а каждое знакоместо  $A_q(I_i)$  содержит код символа  $s_q(I_i)$  и его оценку  $w_q(I_i)$  в диапазоне  $0 \div 1$ . Значения  $1-w_q(I_i)$  будем интерпретировать как вероятности ошибочного распознавания знакоместа  $A_q(I_i)$ .

Пусть поле состоит из одного слова (например, фамилии). Тогда для поля, представимого в виде строки  $S(I_i) = \{s_1(I_i), \dots, s_l(I_i)\}$  и оценок  $\{w_1(I_i), \dots, w_l(I_i)\}$ , определим оценку всего поля:

$$\Theta(S(I_i)) = \min_{1 \leq q \leq l} w_q(I_i). \quad (1)$$

Пусть имеется последовательность распознанных слов с оценками

$$\{(S(I_1), \Theta(S(I_1))), \dots, (S(I_k), \Theta(S(I_k)))\}.$$

Необходимо определить наилучший результат распознавания по нескольким первым словам, минимизируя число выбранных слов. Для этого объединим одинаковые строки в популяции для организации отбора и ранжирования популяций [3–5]. Каждая популяция  $Pp$  состоит из слова  $S(Pp)$ , оценок знакомест  $\{w_1(Pp), \dots, w_l(Pp)\}$ , количества строк  $|Pp|$ , попавших в популяцию, и оценки популяции  $\Theta(Pp)$ .

При первом обнаружении поля создается популяция из одного слова с оценкой,

равной оценке этого слова. При дальнейшем добавлении строки  $S(I_i)$  оценка  $w_q(Pp)$  знакоместа  $A_q$  вычисляется следующим образом:

$$1 - (1 - w_q(Pp)) (1 - w'_q(I_i)),$$

где  $w'_q(I_i)$  – оценка добавляемой строки, а  $w_q(Pp)$  – оценка знакоместа популяции. После очередного шага  $t$  происходит оценка состояния популяций путем сравнения каждого с заранее заданными оценками. При нахождении хотя бы одной популяции, удовлетворяющей условиям, процедура рассмотрения новых строк прекращается, и в качестве наилучшего результата берется эта самая популяция. В случае наличия нескольких таких популяций, берется популяция с наилучшей оценкой.

Подход может быть усовершенствован, если оценка поля вычисляется следующим образом:

$$\Theta(S(I_i)) = \prod_{q=1}^l w_q(I_i); \quad (2)$$

или по формуле

$$\Theta(S(I_i)) = 1 - \prod_{q=1}^l (1 - w_q(I_i)). \quad (3)$$

Для полей, состоящих из нескольких слов (например, место рождения), предлагается сделать разбиение на отдельные слова с последующей конкатенацией выбранных «наилучших» вариантов слов в качестве результата.

Данный алгоритм был проверен на некотором тестовом наборе данных с хорошим и средним качеством видеосъемки паспортов РФ, т.е. видеопоследовательности могли иметь незначительные дефекты (блики, смазывания, дефокусировка) [1]. Набор состоял из 67 видеопоследовательностей объемом от 17 до 100 кадров. Анализировалось распознавание семи полей (ФИО, номер и серия паспорта, дата и место рождения), одно поле (место рождения) состояло из нескольких слов, остальные поля – из одного слова.

Для оценки предложенного алгоритма интеграции использовались две характеристики:

- точность, определяемая как доля результатов распознавания, совпадающих с истинным значением поля,
- ускорение, определяемое как среднее значение количества слов, необходимых для интеграции, отнесенного к количеству кадров видеопоследовательности.

Характеристики алгоритма, полученные на данном наборе, сведены в табл. 1–3.

Таблица 1

Характеристики алгоритма с использованием формулы (1)

| Поле           | Точность | Ускорение | Среднее число кадров |
|----------------|----------|-----------|----------------------|
| Фамилия        | 0,99     | 0,14      | 5,47                 |
| Имя            | 0,99     | 0,1       | 4,59                 |
| Отчество       | 0,99     | 0,11      | 4,91                 |
| Серия паспорта | 0,97     | 0,08      | 3,78                 |
| Номер паспорта | 0,99     | 0,08      | 3,68                 |
| Дата рождения  | 0,96     | 0,08      | 4,34                 |
| Место рождения | 0,91     | 0,22      | 8,87                 |

Таблица 2

Характеристики алгоритма с использованием формулы (2)

| Поле           | Точность | Ускорение | Среднее число кадров |
|----------------|----------|-----------|----------------------|
| Фамилия        | 0,93     | 0,18      | 8,66                 |
| Имя            | 0,99     | 0,1       | 4,62                 |
| Отчество       | 0,99     | 0,11      | 5,16                 |
| Серия паспорта | 0,97     | 0,08      | 3,72                 |
| Номер паспорта | 0,99     | 0,08      | 3,72                 |
| Дата рождения  | 0,96     | 0,08      | 3,31                 |
| Место рождения | 0,84     | 0,23      | 8,12                 |

Таблица 3

Характеристики алгоритма с использованием формулы (3)

| Поле           | Точность | Ускорение | Среднее число кадров |
|----------------|----------|-----------|----------------------|
| Фамилия        | 0,97     | 0,16      | 7,34                 |
| Имя            | 1        | 0,12      | 5,47                 |
| Отчество       | 1        | 0,11      | 5,34                 |
| Серия паспорта | 0,97     | 0,13      | 5,84                 |
| Номер паспорта | 0,91     | 0,08      | 4,09                 |
| Дата рождения  | 0,97     | 0,08      | 3,16                 |
| Место рождения | 0,85     | 0,24      | 9,41                 |

### Заключение

Отметим, что в основном ошибки алгоритма интеграции объясняются случаями отсутствия правильных результатов распознавания, а также ошибками поиска границ поля. Также отметим, что, несмотря на низкие средние значения ускорения, реальные значения  $r/(z_2 - z_1)$ , для некоторых видеопоследовательностей точность близка к 1.

Что касается среднего числа кадров, необходимого для распознавания, то для поля

«место рождения» получаем, что в среднем достаточно 8–9 кадров, для остальных полей – 3–5. Поле «место рождения», как правило, состоит из нескольких слов, и для каждого из них происходит выбор наилучшего. Поэтому полученное для него значение объясняется тем, что при отборе каждого слова число кадров, необходимое для распознавания и достижения при этом заданного уровня точности, может быть различным, и, следовательно, нужно взять наибольшее.

Таким образом, предложенный алгоритм интеграции позволяет получить высокую точность распознавания полей при существенном ограничении качества исходных данных для распознавания.

*Работа была выполнена при поддержке РФФИ (проект № 16–29–09508).*

#### Список литературы

1. Арлазаров В.В., Жуковский А.Е., Кривцов В.Е., Николаев Д.П., Полевой Д.В. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видеокамер для распознавания документов // Информацион-

ные технологии и вычислительные системы. – 2014. – №3. – С. 71–78.

2. Арлазаров В.Л., Марченко А.А., Шоломов Д.Л. Накопительные контексты в задаче распознавания // Труды ИСА РАН. – 2014. – Т. 64; № 4. – С. 64–72.

3. Bangdiwala I.S., Monroe R.J. Some sequential procedures for ordering populations according to means, variances and regression coefficients // Institute of statistics Mimeo Series. № 202, 1958. – P. 78.

4. Shanti S. Gupta, S. Panchapakesan. On Sequential ranking and selection procedures. Technical Report #88–32C // Department of Statistic Purdue University, 1988. – P. 30.

5. Jayanti Chotai. Selection and rankong procedures based on likelihood rations // Department of Mathematical Statistics University of Umeå, 1979. – P. 29.